

Research Article

Preschoolers Benefit From Visually Salient Speech Cues

Kaylah Lalonde^a and Rachael Frush Holt^a

Purpose: This study explored visual speech influence in preschoolers using 3 developmentally appropriate tasks that vary in perceptual difficulty and task demands. They also examined developmental differences in the ability to use visually salient speech cues and visual phonological knowledge.

Method: Twelve adults and 27 typically developing 3- and 4-year-old children completed 3 audiovisual (AV) speech integration tasks: matching, discrimination, and recognition. The authors compared AV benefit for visually salient and less visually salient speech discrimination contrasts and assessed the visual saliency of consonant confusions in auditory-only and AV word recognition.

Results: Four-year-olds and adults demonstrated visual influence on all measures. Three-year-olds demonstrated visual influence on speech discrimination and recognition measures. All groups demonstrated greater AV benefit for the visually salient discrimination contrasts. AV recognition benefit in 4-year-olds and adults depended on the visual saliency of speech sounds.

Conclusions: Preschoolers can demonstrate AV speech integration. Their AV benefit results from efficient use of visually salient speech cues. Four-year-olds, but not 3-year-olds, used visual phonological knowledge to take advantage of visually salient speech cues, suggesting possible developmental differences in the mechanisms of AV benefit.

Speech is inherently multimodal (Munhall & Vatikiotis-Bateson, 2004; Yehia, Rubin, & Vatikiotis-Bateson, 1998), and audiovisual (AV) integration is an important part of speech perception (Rosenblum, 2005). In adulthood, visual information enhances speech perception beyond what is possible with auditory information alone, particularly when the auditory signal is degraded by noise (e.g., Sumbly & Pollack, 1954) or hearing loss (e.g., Erber, 1969). The AV signal can provide a 6- to 15-dB advantage in speech-recognition-in-noise thresholds (MacLeod & Summerfield, 1987) and approximately 30% to 50% advantage in accuracy of speech recognition over the auditory-only signal (Binnie, Montgomery, & Jackson, 1974; Holt, Kirk, & Hay-McCutcheon, 2011; Ross et al., 2011; Ross, Saint-Amour, Leavitt, Javitt, & Foxe, 2007; Wightman, Kistler, & Brungart, 2006). Visual speech and AV integration also play a role in the early development of phonetic categories (Teinonen, Aslin, Alku, & Csibra, 2008). Given that AV speech is ecologically valid, plays a role in speech

perception development, and is especially important for individuals with hearing loss, it is theoretically and clinically important to characterize the trajectory of AV speech integration development.

Literature on AV speech integration development suggests that infants integrate auditory and visual speech cues, but young children do not. Within the first year of life, infants are sensitive to the correspondence between the auditory and visual speech signals (e.g., Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2003), show evidence of McGurk-like AV illusion percepts (Burnham & Dodd, 2004; Desjardins & Werker, 2004; Rosenblum, Schmuckler, & Johnson, 1997), and use visual speech cues to improve speech perception in noise (Hollich, Newman, & Jusczyk, 2005). There are mixed findings regarding AV integration in children, with the literature emphasizing protracted development. Although children 3 years and older evidence AV benefit (Dodd, 1977; Holt et al., 2011; Ross et al., 2011; Wightman et al., 2006), younger children benefit less than adults (Ross et al., 2011). And in some cases, young children show no AV benefit whatsoever (Jerger, Damian, Spence, Tye-Murray, & Abdi, 2009; Wightman et al., 2006). In addition, whereas adults typically fuse discrepant auditory and visual (McGurk) stimuli or report the visual percept, children are much less likely to fuse the auditory and visual stimuli and typically report the auditory percept (Desjardins, Rogers, & Werker, 1997; Massaro,

^aIndiana University, Bloomington

Correspondence to Kaylah Lalonde, who is now at the University of Washington: klalonde@uw.edu

Rachael Frush Holt is now at The Ohio State University, Columbus

Editor: Nancy Tye-Murray

Associate Editor: Karen Kirk

Received December 28, 2013

Revision received June 20, 2014

Accepted September 23, 2014

DOI: 10.1044/2014_JSLHR-H-13-0343

Disclosure: The authors have declared that no competing interests existed at the time of publication.

1984; Massaro, Thompson, Barron, & Laren, 1986; McGurk & MacDonald, 1976).

Mixed findings regarding AV integration in young children have led some to posit that AV speech integration development follows a U-shaped trajectory (Jerger et al., 2009), which has been explained using *dynamic systems theory* (Smith & Thelen, 2003). According to this theory, development does not proceed monotonically; rather, there are periods of disorganization and behavioral regression associated with the emergence of new states of organization (Gershkoff-Stowe & Thelen, 2004; Markovitch & Lewkowicz, 2004). From this perspective, the plateau of the function (when visual speech lacks influence) reflects a period of transition rather than a “loss” of skill. While phonetic representations reorganize, visual phonetic information is thought to be harder to access (Jerger et al., 2009). Two periods of development are commonly associated with phonological reorganization: (a) the vocabulary growth spurt that begins around 18 months of age and extends into middle childhood (Walley, Metsala, & Garlock, 2003) and (b) early literacy training that extends between 6 and 9 years of age (Anthony & Francis, 2005; Morais, Bertelson, Cary, & Algeria, 1986). Most importantly, dynamic systems theory posits that multiple factors—including perceptual skills, phonological representations/knowledge, and general attention resources (Jerger et al., 2009)—interact to account for the U-shaped trajectory. This study emphasizes one factor that might play an important role in the mixed findings regarding visual speech influence in young children (and thus the U-shaped developmental trajectory): differences in the non-sensory cognitive and linguistic (phonological) requirements of tasks used to assess children’s AV integration across studies.

In contrast to infant AV speech integration studies, which typically use indirect tasks that require only automatic responses (see Gerken, 2002, for a review of infant methods), researchers usually test children’s AV integration using more cognitively taxing procedures—the same tasks used with adults. Children are required to make overt responses, including verbal responses (Boothroyd, Eisenberg, & Martinez, 2010; Dodd, 1977; Holt et al., 2011; Jerger et al., 2009; Massaro et al., 1986; McGurk & MacDonald, 1976; Ross et al., 2011; Tremblay et al., 2007) and other motor responses (Desjardins et al., 1997; Hnath-Chisolm, Laipply, & Boothroyd, 1998; Massaro et al., 1986; Sekiyama & Burnham, 2008; Wightman et al., 2006). Because the same tasks are used to assess children and adults, many of the developmental results—in particular, findings of reduced visual influence and AV advantage in young children on some measures—might reflect differences in non-sensory processing efficiency and susceptibility to task demands, rather than age-related changes in AV integration (Allen & Wightman, 1992; Boothroyd, 1991; Wightman, Allen, Dolan, Kistler, & Jamieson, 1989).

When researchers carefully limit non-sensory task demands, developmental differences in AV integration decrease (e.g., Desjardins et al., 1997; Jerger et al., 2009). For example, Jerger and colleagues (2009) limited non-sensory

cognitive and linguistic task demands by using an indirect measure of AV integration. Indirect tasks do not require conscious retrieval, so indirect tasks can be completed with less detailed visual phonological representations than direct tasks (Jerger et al., 2009; Yoshida, Fennell, Swingley, & Werker, 2009). Four-year-old children demonstrated visual influence on Jerger and colleagues’ indirect measure (the multimodal picture naming game) despite demonstrating limited visual influence on direct measures (e.g., McGurk and AV-benefit experiments), suggesting that negative findings on direct measures reflect task demand effects.

Desjardins and colleagues (1997) also carefully limited non-sensory task demands. In order to make the experiment more engaging than previous McGurk experiments with preschool-age participants and to avoid requiring verbal responses, Desjardins and colleagues (1997) had teachers teach the names of toys during story time each day during the month prior to testing. The toy names corresponded to the syllables in the experiment, so the 3- to 5-year-old children were simply required to choose the toy that was named. They also provided verbal and tangible reinforcement for continued participation. Children in this study evidenced greater visual influence than those in other developmental studies using the McGurk effect, suggesting that procedural modifications aimed at decreasing non-sensory cognitive demands can decrease developmental differences between adults’ and children’s AV integration.

AV speech integration measures differ in the degree to which they require participants to access visual (or multimodal) phonological representations and/or phonological knowledge. Participants can complete some AV speech-integration tasks by relying on the salient physical features of AV stimuli (i.e., AV speech detection and discrimination benefit; Bernstein, Auer, & Takayanagi, 2004; Eramudugolla, Henderson, & Mattingly, 2011; Grant & Seitz, 2000; Schwartz, Berthommier, & Savariaux, 2004). Other AV integration tasks require participants to access visual (or multimodal) phonological and lexical representations (Eskelund, Tuomainen, & Anderson, 2011; Schwartz et al., 2004; Tuomainen, Andersen, Tiippana, & Sams, 2005; Tye-Murray, Sommers, & Spehar, 2007; Vatakis, Ghazanfar, & Spence, 2008). For example, it is not possible to successfully recognize AV speech based on only the physical features of the stimulus (Erber, 1982). Instead, recognition—by definition—requires that one access phonological and/or lexical knowledge. Children demonstrate mature AV integration earlier in development when tested using tasks that do not require them to access phonological representations than when tested using tasks that require phonetic decisions (Tremblay et al., 2007), suggesting that the mixed findings regarding preschoolers’ AV integration might reflect differences in the phonological requirements of the tasks used to assess them.

In summary, the mixed findings regarding preschoolers’ AV integration might reflect differences in the cognitive and linguistic demands of tasks used to assess their AV speech integration. One goal of this study was to investigate the mixed findings regarding visual speech influence in preschoolers. To that end, we tested preschool

children using three developmentally appropriate tasks that vary in perceptual difficulty and processing demands to allow preschoolers a better opportunity to demonstrate AV integration. These include tasks that assess sensitivity to congruency of the auditory and visual stimuli as well as AV benefit to speech discrimination and speech recognition. The other goal was to investigate developmental differences in the ability to use visually salient speech cues and visual phonological knowledge. Specifically, we investigated whether the speech contrasts that are visually salient for adults (e.g., Braidá, 1991) and older children (ages 9–15 years; Erber, 1972) are visually salient for preschoolers. In particular, we examined whether AV speech perception benefit in children (relative to auditory-only speech) reflects sensitivity to salient visual speech cues. Visual saliency is a basic property of the speech stimulus, but AV speech likely activates visual (or multimodal) phonological representations. As noted in previous paragraphs, tasks vary in the degree to which they require access to visual (or multimodal) phonological representations. On tasks that only require perceptual sensitivity to the physical features of a stimulus (i.e., discrimination), we expect preschoolers to demonstrate sensitivity to the visual saliency of a speech stimulus. However, tasks that require preschoolers to access visual (or multimodal) phonological representations (i.e., recognition) force preschoolers to rely on their incompletely developed visual phonological knowledge to take advantage of visual speech saliency (Metsala & Walley, 1988; Walley, 2004). We expect to observe developmental differences in the ability to use visual phonological knowledge to take advantage of visual speech saliency.

To address the goals of the investigation, 3- and 4-year-olds completed developmentally appropriate AV speech matching, discrimination, and recognition tasks. In the event that preschoolers demonstrated AV benefit in speech sound discrimination, we investigated whether it resulted from efficient use of visually salient speech cues by comparing benefit for visually salient and less visually salient speech contrasts. In the event that preschoolers demonstrated AV benefit in speech recognition, we investigated whether it resulted from efficient use of salient visual speech cues and visual–phonological knowledge by examining the visual saliency of consonant substitution errors in the auditory-only and AV conditions. Adults were tested using the same tasks but with age-appropriate differences in methodology (i.e., a more difficult signal-to-noise ratio [SNR]) to determine whether the pattern of results observed for children was mature.

We hypothesized that young children and adults would demonstrate AV integration on these developmentally appropriate tasks. Specifically, children and adults would match auditory speech to the appropriate visual articulation with greater-than-chance accuracy and perform better in AV conditions than in auditory-only conditions. If participants' AV benefit results from efficient use of visually salient speech cues (as a physical feature), we hypothesized that they would benefit more on the visually salient speech discrimination contrast than on the less visually salient speech

discrimination contrast. If participants used visual phonological knowledge to take advantage of visually salient speech cues, a greater proportion of consonant substitution errors in the word recognition task would be consistent with visual information (within the same viseme category) in the AV condition than in the auditory-only condition.

Method

Participants

Three groups participated in three experiments: 12 adults (10 women, 2 men), 12 four-year-olds (8 girls, 4 boys), and 15 three-year-olds (9 girls, 6 boys). Adults were 18 to 31 years old ($M = 22.33$ years, $SD = 3.89$ years). Four-year-olds' ages varied between 4.03 and 4.96 years ($M = 4.66$ years, $SD = 0.28$ years). Three-year-olds' ages varied between 3.25 and 3.93 years ($M = 3.59$ years, $SD = 0.24$ years). Preschoolers, especially 3-year-olds, required more testing sessions to collect the same amount of data. To reduce the attrition rate among preschoolers, data were not discarded when participants completed only a subset of the experiments. Six of the 3-year-olds completed a subset of the experiments: One completed only the matching experiment, two completed only the discrimination experiment, one completed only discrimination and recognition, and two completed only matching and recognition. The average age of the 3-year-olds who completed each condition was similar ($M = 3.57$ years for matching, $M = 3.59$ years for discrimination, and $M = 3.58$ years for recognition). One 4.25-year-old completed all of the experiments, but the experimenter failed to properly save the discrimination and recognition data. The matching results include data from all 12 four-year-olds. The discrimination and recognition experiments include data from the remaining 11 four-year-olds ($M_{age} = 4.70$ years, $SD = 0.26$ years). All participants had normal speech, language, hearing, and vision and were from American English-speaking homes. The Indiana University Institutional Review Board approved this study. Participant or parental consent was obtained from/for all participants.

Adults were required to pass a pure-tone hearing screening at 20 dB HL at octave intervals from 250 Hz to 8000 Hz. They also completed a visual discrimination task to screen for sufficient visual acuity (with corrective lenses, if needed). It is unclear exactly what facial cues participants might use for perception, and thus it is unknown exactly what level of visual acuity is necessary to benefit from visual speech cues. To get a sense of participants' visual acuity, the experimenter presented one gray and one black-and-white striped Precision Vision Patti Stripes Square Wave Grating Paddle (Precision Vision; La Salle, IL) from a distance of 100 cm and asked the participant to indicate which paddle had stripes. Although the experimenter asked the participant to point to the striped paddle, the experimenter could also rely on the participant's gaze, which is automatically drawn to the striped paddle. To familiarize participants with the task, the experimenter first presented

the paddle with 52-cycles-per-degree stripes. This was followed by one presentation of the 1.04-cycles-per-degree stripes, and five presentations of the 8.4-cycles-per-degree stripes. To pass the screening, the participants had to correctly identify the 8.4-cycles-per-degree striped paddle on four of five trials. In the following sections, we relate this level of visual acuity to the visual stimuli used in each experiment.

Children underwent vision, hearing, speech, and language screenings. They completed the same visual discrimination task as adults. Children's hearing was screened using four-frequency distortion-product otoacoustic emission screening. If children failed the screening (pass criterion of 6 dB SNR at three or more test frequencies in both ears), a pure-tone screening at 20 dB HL at octave intervals from 250 Hz to 8000 Hz was conducted using conditioned-play audiometry. Speech and language were assessed using the Preschool Language Scale–Fourth Edition screener (Zimmerman, Steiner, & Pond, 2002) on a pass/fail basis.

Matching Experiment

The AV matching experiment used procedures adapted from the infant literature (Kuhl & Meltzoff, 1982) in which researchers assess whether infants look longer at a face that matches the vowel they are hearing than another vowel differing in height. This procedure differs from the infant task in several ways. Whereas infant responses are based on preferential looking times, children were required to make a conscious choice and perform an overt motor response. The stimuli in the current experiment were a sequence of syllables rather than the sustained vowels used with infants in this paradigm. Finally, this experiment had 40 trials, rather than one 2-minute trial.

Stimuli

The stimuli for the discrimination experiment were part of a larger set of professional recordings of a 20-year-old White female talker from the Midwest. She was instructed to sound pleasant and natural and to begin and end each utterance with her mouth closed. The stimuli used in the matching experiment were nonsense /ba/ and /bu/ syllables, each uttered four times in a row. Each syllable in the set of stimuli was equated to the same total root-mean-square power. Ten adults identified the stimuli included in the experiment with 100% accuracy in quiet auditory-only and AV conditions. The same group rated each stimulus at a mean of 6 or higher on a scale of 1 (*poor example of the target stimulus*) to 7 (*excellent example of the target stimulus*) in both auditory-only and AV conditions.

Five sets of /ba ba ba ba/ tokens were temporally matched with five sets of /bu bu bu bu/ tokens such that the onset and offset of each /ba/ syllable were within 73 ms of the onset and offset of each /bu/ syllable. Adults judge matched auditory and visual consonants as simultaneous for auditory leads of up to 74 ms and visual leads of up to 131 ms (van Wassenhove, Grant, & Poeppel, 2007). Children require larger differences to detect temporal asynchrony (Hillock-Dunn & Wallace, 2012). Thus, the adults and

children should have temporally binded the auditory signal with both of the visual signals. The matching stimuli were presented in quiet.

Visual screening ensured that peripheral visual acuity was more than sufficient to discriminate the two vowels. From the viewing distance of 40 cm, the head vertically subtended 18.46° of visual angle. The difference in mouth height for /u/ and the mouth height for /a/ corresponded to about 1.1° of visual angle, meaning that participant's visual acuity was sufficient to see differences on the order of one-ninth the difference in height between the two vowels.

Apparatus

The experimenters used E-Prime Version 2.0 software (Psychology Software Tools, 2007) on an Intel desktop computer to present stimuli and record data. The visual stimuli were routed to a 19-in. Elo Touchsystems touch-screen monitor (Elo Touch Solutions; Milpitas, CA); the auditory stimuli were routed through an audiometer to two wall-mounted speakers in a double-walled sound booth, at ±45° relative to the listener. We checked calibration on each day of testing to ensure that target stimuli were presented at 65 dBA at the imagined location of the listener's head.

Task

The task was identical for the adults and children. On each trial, two faces were presented side-by-side, simultaneously articulating a sequence of four syllables. Participants were told to touch the face of the person whose visual articulation matched the auditory speech signal. After each trial, a puzzle piece appeared on the screen to help the participants track their progress and to reinforce their effort. Order of auditory stimulus, side of the screen with the face articulating /ba/, and side of the screen with the correct matching face were all counterbalanced across the 40 trials completed by each participant. This ensured that participants could not develop strategies, such as choosing the left face each time they heard /ba/. An experimenter sat in the booth throughout testing to confirm participants' responses and to keep them attentive to the task.

Discrimination Experiment

The experimenters used the change/no-change procedure (Sussman & Carney, 1989) to assess discrimination of visually salient (/ba/ vs. /ga/) and less visually salient (/ba/ vs. /ma/) speech contrasts in noise. Children completed testing in the auditory-only and AV modalities; adults completed testing in auditory-only, visual-only, and AV modalities. The change/no-change procedure involves presenting standard and comparison speech sound stimuli and asking listeners to use a developmentally appropriate motor response to indicate whether a *change trial* (comparison stimuli differ from standard stimuli) or a *no-change trial* (comparison stimuli are the same as standard stimuli) was presented. This procedure has been used successfully in auditory-only conditions with 2.5- to 10-year-old children

and adults with normal hearing (Holt & Carney, 2005, 2007; Holt & Lalonde, 2012; Lalonde & Holt, 2014; Sussman & Carney, 1989) and with children with hearing loss (Carney et al., 1993; Osberger et al., 1991).

Stimuli

The stimuli were chosen from the same set as those used in the matching experiment. They met the same identification and rating criteria. Ten sets of /ba ba ba ba/ stimuli were used for no-change trials, and 10 sets each of /ba ga ba ga/ and /ba ma ba ma/ stimuli were used for change trials in the visually salient and less visually salient conditions, respectively. Alternating change-trial stimuli were selected because they provide the listener with multiple opportunities to listen for differences between syllables. Alternating stimuli are just as effective as sequential stimuli (e.g., /ba ba ga ga/) for assessing discrimination in children (Holt, 2011).

A speech-shaped noise was created to match the long-term average spectrum of the stimuli. A 30-ms pink noise was modified to match the long-term average spectrum of the concatenated speech files using a 30-band graphic equalizer in Adobe Audition. The spectra of the visually salient and less visually salient stimuli were similar enough that the same speech-shaped noise was used for both contrasts. The level of the noise within each bandpass filter was altered until the level of the noise matched the level of the speech within 3 dB at each frequency. Random samples of the 30-s noise file with the same duration as the video were mixed with the speech stimuli to create stimuli with a SNR of -5 dB for children and -9 dB for adults. This SNR was chosen for 4-year-olds based on work by Holt and Carney (2007). The same SNR was chosen for 3-year-olds after pilot testing.

In the auditory-only condition, participants saw a monochromatic screen while listening to the auditory stimulus. The solid screen color randomly varied from trial to trial to help maintain attention but in no way was related to the auditory stimulus. In AV conditions, the talker's whole head appeared in the middle of the screen. In the visual-only condition, the same visual images were used as in the AV condition, but the speakers were turned off.

Adults and 4-year-olds sat at a distance of 40 cm from the screen. At this distance, the talker's head height subtended 29.42° of visual angle, and the talker's mouth height subtended 2.86° to 5.01° of visual angle. This means that the adults' and 4-year-olds' visual acuity was sufficient to see visual variation on the order of approximately 1/24 the height of the closed mouth and 1/42 the height of the open mouth. At a distance of 100 cm (as with 3-year-olds), the same stimulus heights correspond to 11.99° of visual angle for the head and 1.15° to 2.01° of visual angle for the mouth. This means the 3-year-olds' visual acuity was sufficient to see visual variation on the order of approximately 1/10 the height of the closed mouth and 1/17 the height of the open mouth. (The difference in distance from the screen for 3-year-olds and other participants is discussed in the 3-Year-Olds subsection.)

Apparatus

The apparatus was the same for the matching and discrimination experiments.

Design and Procedure

Adults were tested in a repeated measures design with six factorial combinations of modality and contrast: auditory-only, visual-only, and AV conditions for both the visually salient (/ba/ vs. /ga/) and less visually salient (/ba/ vs. /ma/) contrasts. Children also were tested in a repeated-measures design, but in a subset of the adult conditions: auditory-only and AV conditions for both the visually salient and less visually salient contrasts. Children were not tested in the visual-only condition because we were primarily interested in the enhancement obtained using the visual speech cues (AV benefit relative to the auditory-only) and because previous research shows that young children with normal hearing do very poorly on visual-only speech perception tasks (e.g., Holt et al., 2011; Jerger et al., 2009). We expected the visual-only condition to yield little useful data and to frustrate children. Adults were tested in the visual-only condition to confirm that the /ba-/ga/ contrast was visually salient and that the /ba-/ma/ contrast was less visually salient. Participants always completed testing in all modalities for a contrast before switching to the other contrast, and modality order was the same for both contrasts. Order of modality and contrast were counterbalanced across participants.

Task

The task differed slightly for each group of participants on the basis of their developmental needs. All participants completed 50 trials per discrimination condition.

Adults. The task used with adults was the same as that used by Holt and Carney (2005). During each trial, a string of syllables was presented in a speech-shaped noise at -9 dB SNR. The string of syllables either changed (e.g., /ba ga ba ga/) or remained the same throughout (e.g., /ba ba ba ba/). The words "change" and "no change" appeared on the screen after the stimulus ended, and adults were instructed to touch "change" if they heard a change in the string of syllables and "no change" if all of the syllables were perceived as the same. After each trial, a puzzle piece appeared on the screen, allowing adults to track their progress. They were told that they could take breaks whenever necessary, but none did.

4-year-olds. In order to make testing more child-friendly, some procedural modifications were introduced for testing 4-year-olds. The same methods were used by Holt and Carney (2007). In the place of the words "change" and "no change," two rows of pictures appeared on the screen. The top row represented a change response and consisted of two sets of two alternating pictures; the bottom row represented a no-change response and consisted of a row of four identical pictures. Children were instructed to touch the row of pictures corresponding to what they heard.

For 4-year-olds, a training phase was also introduced. The training phase consisted of 30 auditory-only trials in quiet. Feedback was provided after each trial, in the form of smiling yellow faces and frowning red faces. Children were required to choose the correct answer on 18 of the last 20 training trials in order to advance to the testing phase. All of the 4-year-old children passed this criterion. Training was completed at the start of each testing day and whenever a new contrast was introduced.

During the testing phase, stimuli were presented in a speech-shaped noise at -5 dB SNR. No feedback was provided on test trials; rather, a puzzle piece appeared on the screen to help keep track of progress and to provide reinforcement. These puzzles had fewer pieces than those shown to adults (10 vs. 30) because in previous work (Holt & Carney, 2007; Holt & Lalonde, 2012; Lalonde & Holt, 2014), we found that children were reinforced by the puzzle activity and liked guessing what picture the puzzle was forming. Children also received noncontingent verbal praise. They received frequent breaks during which they played games and ate snacks. An experimenter sat in the testing booth with the child throughout the experiment and instructed whenever necessary to keep the child attentive to the task. Children were encouraged to look at the screen and were cued to listen and look at the start of each trial.

3-year-olds. The toddler change/no-change procedure (Holt & Lalonde, 2012) was used to test 3-year-olds. This procedure includes a few further modifications meant to make it easier for younger children. Rather than sitting in a chair before a touch-screen monitor, children were tested while sitting or standing on a mat on the floor. This capitalized on their natural desire to move and necessitated a greater distance between them and the monitor than for the 4-year-olds or the adults. They stood or sat on a star on a mat facing two response spaces to the front-left and front-right of the star. The response space on the left contained a row of four identical pictures of ducks (no-change space); the response space on the right contained a row of two sets of alternating pictures of ducks and cows (change space). Rather than touching a screen in response to the stimulus, children jumped to or touched the appropriate response space.

Before training, a 1- to 2-minute live-voice teaching phase was used to help 3-year-old children pair animal sounds with the appropriate pictures of animals on the response mat. Children were required to complete five correct trials in a row to move on to the training phase. The training phase was the same for 3- and 4-year-olds, except that a more lenient training criterion was used for the younger children. Consistent with similar studies (Holt & Lalonde, 2012; Lalonde & Holt, 2014; Trehub, Schneider, & Henderson, 1995), 3-year-olds were required to complete five trials in a row correctly to move on to the testing phase (rather than 18 of the last 20 training trials). Additional forms of reinforcement were integrated into the testing phase. Specifically, an animated reinforcer appeared on the screen following correct responses, and 3-year-olds received tangible reinforcement (listening tickets and snacks) for appropriate listening behavior (being quiet, looking at the screen). Note

that tangible reinforcement was not contingent upon correct responses to each trial.

Recognition Experiment

An AV version of the *Lexical Neighborhood Test* (LNT; Kirk, Pisoni, & Osberger, 1995) was used to assess open-set auditory-only and AV word recognition in noise.

Stimuli

The LNT consists of two lists of 50 monosyllabic words within the productive vocabulary of 3- to 5-year-old children. Half of the words on each list are lexically easy (high familiarity, low neighborhood density), and half are lexically hard (low familiarity, high neighborhood density), according to Luce and Pisoni's (1998) neighborhood activation model of spoken word recognition. The stimuli were professional AV recordings of a professional female announcer. The recording procedures have been described previously (Holt et al., 2011). These stimuli were presented in a steady-state noise with the same long-term average spectrum as the target stimuli. In AV conditions, the talker's whole head appeared in the middle of the screen. The auditory-only condition was identical to the AV condition, except that the screen was turned off.

Participants sat approximately 40 cm from the monitor. From this distance, the head height and mouth height subtended 21.24° and 2.15° of visual angle, respectively. This means that the participants tested had sufficient visual acuity to see visual variation on the order of approximately 1/18 the size of the closed mouth.

Apparatus

Dedicated software was used to present the recognition stimuli. The tester recorded responses on paper.

Design and Procedure

All participants were tested in two conditions: auditory-only and AV. Order of test modality was counter-balanced across participants, but the two lists were always presented in the same order (with List 1 used in the first condition). In this way, half of the participants were tested on List 1 in auditory-only conditions, and half were tested on List 1 in AV conditions.

Task

Fifty monosyllabic LNT words per modality were presented individually in a random order. Participants were instructed to listen to the word in the noise and to repeat it aloud. They were told that they might not hear the whole word and should guess, even if they were not sure. The experimenter phonetically transcribed responses, asking for clarification whenever necessary. Nonword scores were accepted because we planned to carry out phoneme-level analyses.

The recognition task used the same SNRs as the discrimination task. Children were encouraged to look at the screen during the AV trials and were cued to listen and

look at the start of each trial. The animated reinforcers and puzzles were not used in the recognition task. Children received verbal reinforcement (praise) for attempting to respond, regardless of whether they responded correctly, and tangible reinforcement (listening tickets, snacks) for appropriate listening behaviors (sitting quietly, looking at the screen).

Overall Procedure

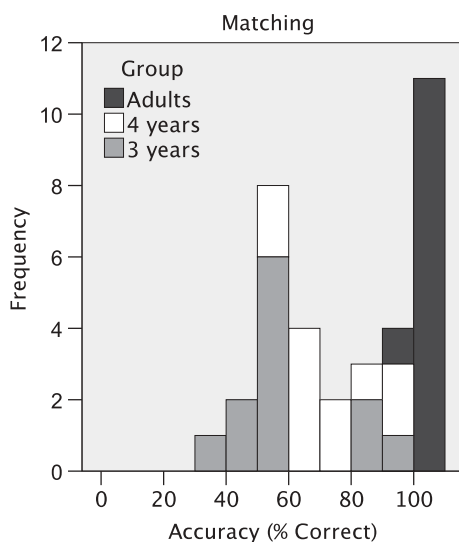
The order of the experiments was counterbalanced across participants in each age group. Order of testing modality was held constant across experiments. In other words, if a participant was tested in AV conditions first in the discrimination experiment, she or he was also tested in AV conditions first in the recognition experiment.

Analysis and Results

Matching Experiment

The dependent variable in the matching experiment was accuracy. The results are displayed in the frequency histogram in Figure 1 for adults (dark gray), 4-year-olds (white), and 3-year-olds (light gray). As expected, adults performed at ceiling on this task ($M = 99.79\%$, $SD = 0.72\%$); only one adult made a single error. All of the 4-year-olds performed better than chance, but performance varied from near chance to almost ceiling levels ($M = 72.20\%$, $SD = 14.57\%$). Mean performance among 3-year-olds was close to chance ($M = 59.79\%$, $SD = 18.27\%$). Of the twelve 3-year-olds tested, three performed below chance (below 50%), six performed barely above chance (51% to 60%), and three had higher accuracy (82% to 95%). Two of the three children with higher accuracy scores were nearly 4 years

Figure 1. Distribution of accuracy scores for audiovisual speech matching by 12 adults (dark gray), 12 four-year-olds (white), and 12 three-year-olds (light gray).



of age (3.9 years), but the other was one of the youngest participants tested (3.3 years).

The accuracy data were transformed to rationalized arcsine units (RAUs; Studebaker, 1985) to stabilize error variance prior to statistical analysis. The RAU data were analyzed using a one-way analysis of variance (ANOVA) with age group as the independent variable. There was a significant effect of group, $F(2, 33) = 45.979$, $p < .001$. Post hoc analyses with Bonferroni corrections for multiple comparisons revealed that the age effect was driven by significantly better performance by adults than by 4-year-olds, $p < .001$, or 3-year-olds, $p < .001$. There were no significant differences in performance between the 3- and 4-year-old groups, $p = .147$. However, one-sample t tests indicated that the 4-year-old group performed better than chance, $t(11) = 4.700$, $p = .001$, but the 3-year-old group did not, $t(11) = 1.851$, $p = .091$.

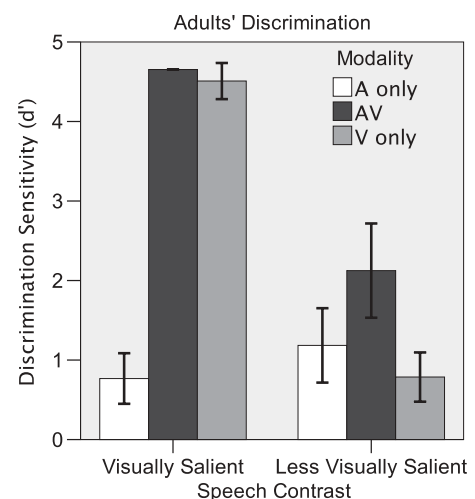
Discrimination Experiment

The dependent variable in the discrimination experiment was d' , a bias-free measure of sensitivity. Due to procedural differences in the ways that children and adults were tested (especially differences in SNR), adult and child data were analyzed separately.

Adults

Figure 2 displays the results of the discrimination experiment performed with adults. A 3 (modality: auditory-only, visual-only, AV) \times 2 (contrast: visually salient, less visually salient) repeated-measures ANOVA was used to evaluate the adult data. The effects of modality, $F(2, 22) = 98.883$, $p < .001$, and contrast, $F(2, 22) = 118.8$, $p < .001$, were significant. There was a significant Modality \times Contrast interaction, $F(2, 22) = 122.6$, $p < .001$. Post hoc paired-samples

Figure 2. Mean performance (± 1 SD) on visually salient (/ba/ vs. /ga/) and less visually salient (/ba/ vs. /ma/) speech contrasts by 12 adults in auditory-only (A only; white), visual-only (V only; light gray), and audiovisual (AV; dark gray) conditions.



t tests with Bonferroni corrections for multiple comparisons were conducted to investigate the effect of contrast in each modality. Performance in the auditory-only condition did not differ significantly between the two contrasts, $t(11) = -1.564$, $p = .164$, suggesting that the contrasts were matched for difficulty in the auditory-only condition. In the visual-only and AV conditions, adults performed much better on the visually salient contrast than on the less visually salient contrast, $t(11) = 23.007$, $p < .001$, and $t(11) = 9.393$, $p < .001$, respectively. These results confirm that the visually salient place contrast was more visually salient than the less visually salient manner contrast and that visual saliency contributes to AV integration benefit in adults.

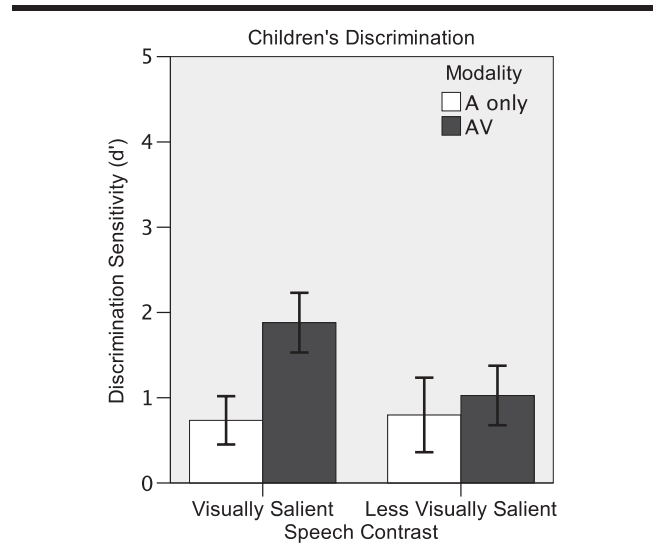
To further investigate the Modality \times Speech Contrast interaction, two separate one-way ANOVAs were performed on the visually salient and less visually salient contrasts. There was a significant effect of modality in both the visually salient, $F(2, 22) = 444.3$, $p < .001$, and less visually salient, $F(2, 22) = 12.27$, $p < .001$, conditions. Post hoc analyses with Bonferroni corrections for multiple comparisons were performed to investigate this effect. Consistent with our assertion that the place contrast was visually salient, adults performed better in the AV and visual-only conditions than in the auditory-only condition on the visually salient contrast, $p < .001$ (mean AV benefit relative to the auditory-only condition = d' difference of 3.88). Adults performed at ceiling in the visual-only and AV conditions, so performance did not differ between these two conditions.

In the less visually salient condition, post hoc analyses with Bonferroni corrections for multiple comparisons indicated that adults performed better in the AV condition than in the auditory-only condition, $p = .03$, or in the visual-only condition, $p = .001$. The AV benefit observed in the less visually salient condition is a bit surprising, given the intuition that listeners rely primarily on the auditory signal for manner of articulation cues (Grant, Walden, & Seitz, 1998). There is clearly some visually salient information available in the stimuli. In fact, adults performed at greater-than-chance levels in the visual-only condition, $t(11) = 5.588$, $p < .001$. Analysis of the stimulus characteristics revealed that the /ma/ syllables were longer than the /ba/ syllables (mean difference = 100 ms). It appears that adults were able to rely on this duration cue to discriminate visual /ba/ and visual /ma/ and to derive AV benefit. Figure 2 shows that this benefit (mean d' difference = 0.94) was not as great as that for the more visually salient condition, supporting the classification of these two contrasts as “visually salient” and “less visually salient.”

Children

The child discrimination results are shown in Figure 3. A 2 (age: 3 years, 4 years) \times 2 (modality: auditory-only, AV) \times 2 (contrast: visually salient, less visually salient) mixed ANOVA was used to analyze the child discrimination data. There was no effect of age, and no variables interacted with age, so results were collapsed across the two age groups. There were significant effects of contrast, $F(1, 21) = 7.675$, $p = .011$, and modality, $F(1, 21) = 23.392$, $p < .001$, and a

Figure 3. Mean performance (± 1 SD) on visually salient (/ba/ vs. /ga/) and less visually salient (/ba/ vs. /ma/) speech contrasts by 23 children in auditory-only (A only; white) and audiovisual (AV; dark gray) conditions.



significant Contrast \times Modality interaction, $F(1, 21) = 18.833$, $p < .001$. A series of post hoc paired-samples *t* tests was conducted to investigate this interaction. As with adults, performance in the auditory-only condition did not differ significantly between the two contrasts, $t(22) = -0.338$, $p = .738$, suggesting that the auditory contrasts were matched for difficulty in children, too. In the AV condition, children performed better on the visually salient contrast than on the less visually salient contrast, $t(22) = 5.310$, $p < .001$. Children performed better in the AV condition than in the auditory-only condition on the visually salient contrast, $t(22) = 6.547$, $p < .001$, suggesting AV benefit to speech discrimination. However, children did not show the same AV benefit for the less visually salient contrast, $t(22) = 1.338$, $p = .195$. This difference between the adult and child data suggests that children might not be able to use subtler visual cues—such as the duration cue in the less visually salient contrast—to derive AV benefit to speech discrimination. These results demonstrate that visual saliency contributes to AV integration benefit in children.

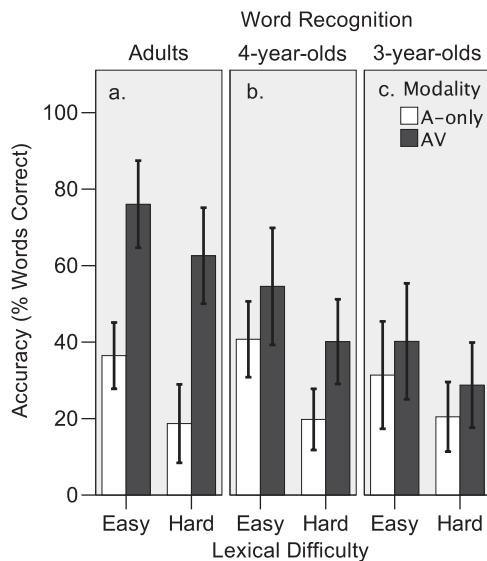
Recognition Experiment

The recognition results were scored by proportion of words correct and proportion of phonemes correct. Accuracy data are shown in Figures 4 and 5 for words and phonemes, respectively. The data were again RAU-transformed before analysis. Due to procedural differences in the way children and adults were tested (in particular, differences in SNR), adult and child data were analyzed separately.

Adults

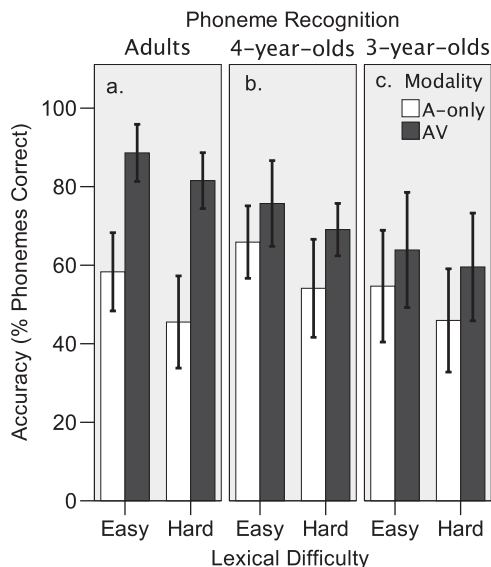
Two 2 (modality: auditory-only and AV) \times 2 (difficulty: lexically hard, lexically easy) repeated-measures ANOVAs were conducted to analyze the adult data; one on the

Figure 4. Mean accuracy (± 1 SD) of lexically easy and lexically hard word recognition in auditory-only (A-only; white) and audiovisual (AV; gray) modalities, by 12 adults (Panel a), 11 four-year-olds (Panel b), and 12 three-year-olds (Panel c). This figure displays the percentage of words correctly recognized.



word-correct data and one on the phoneme-correct data. As shown in Figure 4a, adults correctly recognized more lexically easy words than lexically hard words, $F(1, 11) = 110.2, p < .001$, and recognized more words in the AV modality than in the auditory-only modality, $F(1, 11) = 35.569, p < .001$. There was no significant Modality \times Lexical

Figure 5. Mean accuracy (± 1 SD) of lexically easy and lexically hard word recognition in auditory-only (A-only; white) and audiovisual (AV; gray) modalities, by 12 adults (Panel a), 11 four-year-olds (Panel b), and 12 three-year-olds (Panel c). This figure displays the percentage of phonemes correctly recognized.



Difficulty interaction. Figure 5a shows that the same pattern emerged at the phoneme level. There was an effect of lexical difficulty, $F(1, 11) = 22.252, p < .001$, and modality, $F(1, 11) = 112.3, p < .001$. These modality effects demonstrate AV benefit to speech recognition in adults. The lack of interaction between modality and lexical difficulty suggests that the degree of visual enhancement was independent of lexical difficulty.

Children

Two 2 (age: 3 years, 4 years) \times 2 (modality: auditory-only and AV) \times 2 (difficulty: lexically hard, lexically easy) mixed ANOVAs were conducted to analyze the child data. The dependent variable in the first analysis was word-recognition accuracy. The results are shown in Figure 4b for 4-year-olds and in Figure 4c for 3-year-olds. There were significant effects of modality, $F(1, 21) = 44.031, p < .001$, and lexical difficulty, $F(1, 21) = 155.809, p < .001$. Like adults, children recognized more AV words ($M = 40.64\%$) than auditory-only words ($M = 28.00\%$), and they recognized more lexically easy words ($M = 41.47\%$) than lexically hard words ($M = 27.17\%$). The effect of age was not quite significant, $F(1, 21) = 4.308, p = .05$, but there was a significant Age \times Lexical Difficulty interaction, $F(1, 21) = 3.916, p = .021$. Independent-samples t tests with age as the independent variable and Bonferroni corrections for multiple comparisons indicated that 4-year-olds recognized more easy words than did 3-year-olds, $t(19) = 2.648, p = .032$. There was no significant difference between the number of hard words that 3- and 4-year-old children recognized, $t(19) = 1.661, p = .226$. Although the average AV benefit was greater in 4-year-olds ($M = 17.10\%$) than in 3-year-olds ($M = 7.87\%$), there was no Group \times Modality interaction. The differences in AV-benefit between 3- and 4-year-old participants were not significant. Overall, and with regard to the issue of AV integration benefit, the word-recognition data showed an adultlike pattern. Preschoolers demonstrated AV benefit to speech recognition, and the degree of visual influence was independent of lexical difficulty.

The 2 (age) \times 2 (modality) \times 2 (difficulty) mixed ANOVA at the phoneme level demonstrated developmental effects. Results for 4-year-olds and 3-year-olds are shown in Figures 5b and 5c, respectively. Four-year-old children performed significantly better than 3-year-old children, $F(1, 21) = 5.844, p = .025$. The average AV benefits for 4-year-olds ($M = 12.2\%$) and 3-year-olds ($M = 11.3\%$) were similar. There were no developmental differences in the pattern of results (no interaction between age and other factors), so results for the two age groups were analyzed together. Preschoolers recognized more phonemes in lexically easy words than in lexically hard words, $F(1, 21) = 41.171, p < .001$, and more phonemes in the AV modality than in the auditory-only modality, $F(1, 21) = 36.431, p < .001$. Like adults, preschoolers demonstrated AV benefit to speech recognition, and the degree of AV benefit was independent of lexical difficulty at both the word level and the phoneme level.

Consonant Confusion Analysis

We created and analyzed confusion matrices to probe consonant substitution errors and explore the role of visual saliency and visual phonological knowledge in adults' and children's AV speech recognition. Several researchers have created confusion matrices based on adults' visual-only (lipreading) consonant substitutions (see Owens & Blazek, 1985, for review). This allowed them to determine which phonemes are visually distinct and which are highly confusable. Highly confusable phonemes were placed in the same viseme category (Fisher, 1968). Visual information cannot reliably distinguish phonemes in the same viseme category. Visually distinct phonemes belong to different viseme categories. They are not often confused with one another in visual-only testing because the visual information reliably distinguishes them.

For the current investigation, we examined consonant substitutions in auditory-only and AV conditions. Each substitution error was classified as either a within-viseme category error or a between-viseme category error, based on Owens and Blazek's (1985) viseme category classifications. The dependent variable in the consonant confusion analysis was the proportion of substitution errors that were within-viseme category substitutions (i.e., the proportion of consonant confusions that were within the same viseme category and thus visually confusable).

Adults

Figure 6a shows the proportion of adult consonant confusions that were within-viseme category substitutions as a function of modality. These data were RAU-transformed and analyzed using a paired-samples *t* test with modality as the independent variable. A greater proportion of substitution

errors were within-viseme category substitutions in the AV modality ($M = 47.00\%$) than in the auditory-only modality ($M = 15.73\%$), $t(11) = 6.045$, $p < .001$.

Children

A similar analysis was conducted to investigate the proportion of child consonant confusions that were within-viseme category substitutions as a function of modality. As for adults, the RAU data were analyzed using paired-samples *t* tests with modality as the independent variable. As shown in Figure 6b, 4-year-old children demonstrated an adultlike pattern of results. A greater proportion of substitution errors were within-viseme category substitutions in the AV modality ($M = 32.18\%$) than in the auditory-only modality ($M = 16.64\%$), $t(10) = 2.929$, $p = .015$. As shown in Figure 6c, the pattern of results was different for 3-year-old children. There was no difference in the proportion of substitution errors that were within-viseme category substitutions in the auditory-only ($M = 26.61\%$) and AV ($M = 30.83\%$) modalities, $t(11) = 0.943$, $p = .366$.

Discussion

The purpose of this study was twofold: (a) to explain the mixed findings regarding visual speech influence in preschoolers and (b) to investigate developmental differences in the ability to use visually salient speech cues and visual phonological knowledge.

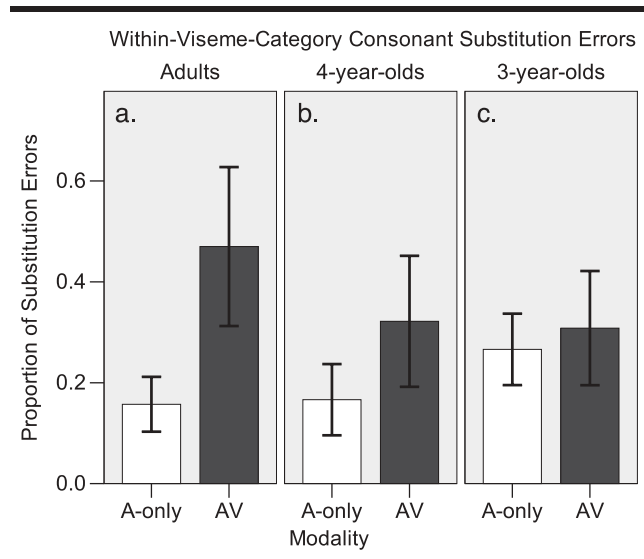
We used developmentally appropriate speech perception tasks varying in difficulty to provide preschoolers a better opportunity to demonstrate visual speech influence. We assessed whether 3- and 4-year-old children can match auditory and visual speech cues and benefit from integrating auditory and visual speech cues at the discrimination and the recognition levels. We also compared AV integration benefit for visually salient and less visually salient speech discrimination contrasts and assessed the visual saliency of consonant confusion errors to determine whether benefit reflects the use of visual saliency and visual-phonological knowledge.

Visual Influence and AV Benefit

Adults and 4-year-olds demonstrated AV integration on the AV matching, discrimination, and recognition tasks. They matched auditory speech to the corresponding articulations with greater-than-chance accuracy, and they discriminated speech sounds and recognized words better in AV conditions than in auditory-only conditions. This suggests that 4-year-old children can demonstrate visual influence on developmentally appropriate direct measures. Three-year-olds demonstrated AV benefit to speech discrimination and recognition, but most scored near chance on the AV matching task.

Previous literature suggests that preschoolers might only be able to demonstrate AV integration on indirect measures, possibly because indirect measures do not require conscious information retrieval (Jerger et al., 2009). Direct

Figure 6. Mean proportion of consonant confusions (± 1 SD) that were within-viseme-category errors by 12 adults, 11 four-year-olds, and 12 three-year-olds, in the auditory-only (A-only; white) and audiovisual (AV; gray) modalities.



measures of preschoolers' AV integration—particularly those using the McGurk paradigm—have demonstrated little or no visual influence. Jerger and colleagues (2009) demonstrated visual influence in preschoolers on an indirect measure and concluded that the negative findings on previous direct measures of AV integration in preschoolers likely reflected task demands rather than age-related effects. The results of the current study indicate that preschoolers can demonstrate AV integration on some direct measures. To the extent that negative findings on previous direct measures represent effects of task demands, the current results suggest that our measures of AV integration were developmentally appropriate for preschoolers. In the sections that follow, we discuss in greater detail why 3-year-olds may have demonstrated mixed findings.

Matching

Although different methods and units of measurement were used, we can compare the matching results from preschoolers and adults in the current investigation to those of infants in previous studies. In the infant studies, the dependent variables were the proportion of looking time spent on the matching face (50% = chance) and the proportion of infants who looked at the matching face more than chance (more than 50% of looking time). Although we measured the proportion of correctly matched trials, we can compare the proportion of children and adults who performed above chance with the proportion of infants who performed above chance. Across studies, 71.88% to 78.13% of the infants looked at the matching face more than 50% of the time (Kuhl & Meltzoff, 1982, 1984; Patterson & Werker, 1999, 2003). The 3-year-olds' data fit within the range reported for infants: 75% of the 3-year-olds performed above chance. In contrast, all of the 4-year-olds and adults performed better than chance.

More research is needed to understand the relationship between the infants' looking-time data and the children's accuracy data. Direct (accuracy) and indirect (looking behavior) measures of AV speech matching need to be collected in the same group of participants. Under the assumption that direct measures require more detailed visual speech representations and visual phonetic knowledge than indirect measures (Jerger et al., 2009), we expect that more than 75% of 3-year-olds would demonstrate visual influence on an indirect matching task.

Other differences between the matching task used in this study and the one used with infants in previous studies may have made this experiment harder than those used with infants. The side of the screen with the face articulating /ba/ and /bu/ varied randomly from trial to trial. The stimuli had more temporal variation than the steady-state vowels used in the infant studies, and there was no familiarization period. Future studies should assess the effect of these methodological differences on 3-year-olds' performance.

We presented stimuli in quiet for this first investigation because pilot testing with 3- and 4-year-olds demonstrated that children did not perform at ceiling in quiet and might have struggled to perform the task in noise. In addition,

whereas participants can rely solely on the auditory information when the discrimination and recognition tasks are easy (making it impossible to observe visual influence), participants must integrate auditory and visual information to perform the matching task, regardless of the level of listening difficulty. Even in quiet, participants must make cross-modal comparisons to determine whether the auditory and visual signals are congruent. Given that adults performed at ceiling on the matching task, future research may include assessing AV matching in noise. Assessing how the difficulty of this task compares with other measures (such as AV benefit to discrimination and recognition) under comparable listening conditions may reveal that adults (and, perhaps, children) rely on phonetic knowledge to match auditory and visual speech.

Discrimination

Although previous research suggests that visual speech improves 6- to 10-year-olds' and adults' phoneme detection in noise on a phoneme-monitoring task (Fort, Spinelli, Savariaux, & Kandel, 2010, 2012), this study is the first we know of to demonstrate AV benefit to speech sound discrimination in preschoolers with normal hearing. In fact, younger children (5-year-olds) fail to demonstrate the same visual phoneme-detection benefit (Fort et al., 2012). Boothroyd and colleagues (2010) completed the only other investigation of AV discrimination benefit in preschoolers. They assessed 2.6- to 6.6-year-old children's discrimination of auditory-only and AV speech pattern contrasts in quiet using the *On-Line Imitative Test of Speech Pattern Contrast Perception* (OlimSpac; Boothroyd, Eisenberg, & Martinez, 2010). There was no effect of modality, likely due to ceiling effects: 27 of the 30 children scored better than 90% in at least one modality.

Although preschoolers in the current study benefited from the AV signal, they were not as sophisticated as adults at using the available visual cues in conjunction with the auditory ones. Adults were able to use manner cues (likely temporal in nature) to enhance their AV discrimination of the less visually salient (/ba/ vs. /ma/) discrimination contrast, whereas preschoolers did not show evidence of using these more subtle visual temporal cues. On the visually salient contrast, adults could rely on the visual signal to achieve ceiling levels of performance. For the less visually salient contrast, adults had to integrate the auditory and visual signals to enhance their AV discrimination: Both auditory-only and visual-only performance on the less visually salient contrast were poorer than AV performance. These adult findings are consistent with early work on the contribution of auditory and visual speech to adults' AV consonant perception. This work showed that the visual signal increases transmission of place information more than other information, such as manner information (Binnie et al., 1974). It is important that the current results partially extend these findings to preschoolers: Young children in the discrimination task evidenced AV benefit for the visually salient contrast but not for the less visually salient contrast. This suggests that even young children use visual speech cues in

a speech discrimination task, but they are not as sophisticated as adults in their ability to use all available cues, particularly those that are more subtle (such as the duration cue in the less visually salient contrast).

Our results conflict with those of Jerger et al. (2009), wherein 4-year-olds demonstrated visual speech influence for the less visually salient voicing cue but not the visually salient place cue on the multimodal picture-naming game. Future work should investigate task-related differences in the effects of visual speech cue saliency on visual influence in young children.

Recognition

The degree of AV benefit to speech recognition in this study is similar to that found in previous studies. Holt et al. (2011) assessed AV benefit to speech recognition in 3- to 5-year-old children using the *AV Lexical Neighborhood Sentence Test* (AVLNST; Kirk et al., 1995) at -2 dB SNR. These young children derived approximately the same degree of AV benefit for sentence recognition ($M = 12.72\%$) as did those children in the current study who were tested at -5 dB SNR for word recognition ($M = 12.77\%$). In both the current study and the previous study using the AVLNST, adults tested at more difficult SNRs demonstrated greater benefit than children. Adults' benefit was also similar for the AVLNST at -7 dB SNR ($M = 36.81\%$) and the current study using the LNT at -9 dB SNR ($M = 42.31\%$ for word correct).

The 4-year-old group demonstrated better speech recognition than the 3-year-old group at the phoneme level. (The age effect just missed significance at the word level.) The developmental differences were the same across test modalities and, therefore, likely reflect cognitive and linguistic processes other than AV speech integration. (True developmental differences in AV integration/visual influence should result in an Age \times Modality interaction.) Similar task-related cognitive-linguistic effects might have been observed on the discrimination task if we had not tested 3-year-olds using an easier version of the task (the toddler change/no-change procedure). This result highlights the advantage of using AV benefit measures to assess visual speech influence in young children: We can better differentiate effects of visual influence from those of task-related variables.

Visual Speech Saliency and Visual Phonological Knowledge in AV Speech Benefit

Previous studies of AV speech integration benefit in young children have primarily focused on demonstrating and quantifying visual enhancement. In this study, we also asked whether the speech contrasts that are salient for adults (e.g., Braidá, 1991) and older children (ages 9–15 years; Erber, 1972) are visually salient for preschoolers and whether preschoolers can use their incompletely developed visual phonological knowledge to take advantage of visual speech saliency (Metsala & Walley, 1988; Walley, 2004).

Visual saliency contributed to the AV speech-discrimination benefit in preschoolers and adults: All age groups demonstrated more AV benefit for visually salient speech contrasts than for less visually salient speech contrasts. This dissociation suggests that AV speech benefit in adults and 3- and 4-year-old children reflects the efficient use of visually salient speech cues. Although it is clear that visual saliency contributes to AV speech discrimination benefits in children and adults, it is unclear whether these visual saliency effects reflect the use of visual phonological knowledge. It is possible that speech discrimination activates visual phonological representations, but it is also possible to successfully discriminate speech sounds on the basis of only the physical features of the stimulus.

At the recognition level, adults' and 4-year-olds' substitution errors were more likely to involve visually confusable phonemes in the AV condition than in the auditory-only condition. AV recognition benefit in 4-year-olds and adults depended on the visual saliency of speech sounds. Unlike in the discrimination task, it is not possible to successfully recognize speech on the basis of only the physical features of the stimulus (Erber, 1982). Adults and 4-year-olds must have used visual (or multimodal) phonological representations and phonological knowledge to achieve AV benefit (i.e., knowledge that the salient visual features are associated with the phoneme/word recognized). This was found despite 4-year-olds having relatively less detailed phonological representations (Metsala & Walley, 1988; Walley, 2004) and immature lipreading skills (Jerger et al., 2009; Massaro et al., 1986; Tye-Murray & Geers, 2001) relative to adults.

Three-year-olds' data demonstrated a discrepancy between the accuracy data and the error patterns. Although 3-year-olds demonstrated AV benefit to recognition accuracy, the visual information did not affect the proportion of visually confusable errors. Three-year-olds' error patterns were not as sophisticated as those of 4-year-olds and adults. The error patterns provided no evidence that 3-year-olds used visual-phonological knowledge to derive AV recognition benefit.

This study highlights the possibility that children at different stages of development might rely on different mechanisms to obtain the same AV benefits. Whereas 3- and 4-year-olds demonstrated the same degree of AV speech recognition benefit, only 4-year-olds' errors suggested that they used visual phonological knowledge to derive that benefit. This is not a novel idea: Jerger et al. (2009) suggested that different mechanisms underlie the benefit observed in 4-year-olds and 10- to 14-year-olds on their multimodal picture naming game. They suggested that older children might rely on robust, detailed visual (or multimodal) phonological representations and phonological knowledge, whereas visual influence in the younger group might reflect enhanced general information processing (such as enhanced attention). In other words, visual speech might increase young children's motivation and attention/orientation to the stimulus, which indirectly affects audiovisual speech integration (Arnold & Hill, 2001).

The assertion that there are multiple mechanisms of AV speech integration and that different mechanisms are more or less active at different ages is supported by behavioral and neurophysiological evidence. There are multiple neural pathways that underlie AV speech perception, and different pathways underlie various AV phenomena (Driver & Noesselt, 2008; Eskelund et al., 2011). Some AV speech integration phenomena result from general perceptual mechanisms, such as reduced temporal uncertainty (Bernstein et al., 2004; Eramudugolla et al., 2011; Grant & Seitz, 2000; Schwartz et al., 2004). Others result from speech-specific mechanisms involved in accessing visual (or multimodal) phonological and lexical representations (Eskelund et al., 2011; Schwartz et al., 2004; Tye-Murray et al., 2007; Tuomainen et al., 2005; Vatakis et al., 2008). General perceptual mechanisms of integration mature faster than speech-specific mechanisms (Tremblay et al., 2007), so it is possible that 3-year-olds used a general perceptual mechanism to obtain AV recognition benefit.

One other potential explanation for these results is that some of the 3-year-olds were not always perfectly intelligible. This is consistent with normal speech intelligibility development (Coplan & Gleason, 1988; Weiss & Lillywhite, 1976). Although this is a potential limitation of the study, other studies have also used verbal response methods to assess speech perception and AV integration in 3-year-olds (e.g., Boothroyd et al., 2010; Holt et al., 2011; Massaro et al., 1986; McGurk & MacDonald, 1976; Tremblay et al., 2007). We made every attempt to minimize intelligibility effects. When a participant's response was unclear, we asked for repetition. If it was a choice between two words, we used definitions and descriptions to determine which word they meant. If there was a question about which of two sounds were being produced, we asked participants to choose from examples, such as, "Is that an /r/ as in 'run' or a /d/ as in 'duck'?"

This study highlights the importance of considering the perceptual processing demands of AV speech-perception-benefit tasks. The discrimination task assessed a lower and earlier developing level of speech perception (the perceptual representation level) than the recognition task (cognitive-linguistic level; Aslin & Smith, 1988). The discrimination task requires complex neural encoding of the multimodal stimulus; the recognition task requires that the listener attach a label (and, sometimes, meaning) to this complex neural code (Aslin & Smith, 1988). Although it was possible to obtain an AV discrimination benefit using only the physical features of the stimulus, participants had to rely on visual phonological skills to obtain an AV recognition benefit. Although we discuss qualitative differences in performance on these tasks, it was not possible to directly compare results across the AV speech tasks. The tasks used different stimuli and responses, the dependent variables of the tasks had different units, and chance level varied from task to task. Current research in our lab aims to more thoroughly explore how AV speech integration benefit development interacts with level of perceptual processing. Specifically, we are using AV benefit tasks that vary along

only one dimension—that is, the level of perceptual processing necessary to complete the task (detection, discrimination, recognition)—to examine whether the perceptual processing demands affects AV integration in children.

Changes in visual speech influence with development have been attributed to changes in the perceptual weighting of auditory and visual speech, age-related improvements in lipreading, transitions from undifferentiated holistic to modality-specific processing, increases in the detail of visual speech representations, and changes in general attention mechanisms (Desjardins et al., 1997; Jerger et al., 2009; Massaro, 1984; Massaro et al., 1986; Sekiyama & Burnham, 2008). We observed that (a) 3- and 4-year-olds can use some salient visual speech cues for discrimination and (b) 4-year-olds have visual speech representations that are adequately specified to use some salient visual speech cues for recognition. However, the current study was not designed to assess the trajectory of visual speech influence development or the mechanisms of visual speech influence changes across development; this study was not longitudinal, and the age range was too small to assess dynamic systems theory (Smith & Thelen, 2003). Rather, we assessed an alternative explanation: that differences in the tasks used to assess integration across studies account for a portion of the mixed findings regarding preschoolers' AV integration and affects our current view of the trajectory of visual speech influence development. Dynamic systems theory posits that multiple interactive factors determine the trajectory of development (Smith & Thelen, 2003); therefore, these two explanations are not necessarily in conflict.

The current results suggest that previous research might have underestimated preschoolers' ability to use visual speech because they were tested using relatively difficult tasks. However, more research—for example, using larger age ranges or longitudinal designs and using both direct and indirect measures of visual influence that assess multiple levels of perceptual processing—is needed to continue to understand the development of visual speech influence.

Conclusions

This study leads to the following conclusions:

1. Four-year-old children can demonstrate AV speech integration when developmentally appropriate direct measures are used. Three-year-old children can demonstrate AV speech integration on some developmentally appropriate direct measures.
2. Three- and 4-year-olds use some salient visual speech cues for AV speech discrimination benefit.
3. Four-year-old children can use visual phonological knowledge to obtain AV recognition benefit.
4. Three-year-old children do not demonstrate the use of reliable visual information to improve speech recognition performance. This might reflect difficulty relying on visual phonological knowledge, reliance on

general perceptual mechanisms of visual influence, or task-related variables such as requiring a verbal response from participants who make systematic articulation errors.

5. Although preschoolers can use salient visual speech cues (such as place of articulation) to derive AV benefit, they might not (be able to) rely on subtler visual cues (such as duration cues) to derive AV benefit.
6. Comparing performance across modalities allows us to assess whether developmental differences were likely caused by task-related variables or age-related differences in visual influence. Developmental differences between 3- and 4-year-olds' performance on the recognition task likely reflect immature cognition and language rather than differences in visual speech influence.
7. More research is needed to understand the relationship between the direct and indirect measures of AV integration, mechanisms of AV benefit, and how AV integration is affected by the level of perceptual processing required by a task.

Acknowledgments

This work was supported by National Institute on Deafness and Other Communication Disorders Predoctoral Training Grant T32 DC00012 (awarded to David B. Pisoni) and by the Ronald E. McNair Research Foundation. We are grateful for the expertise of Rowan Candy, who provided an age-appropriate measure of visual acuity. We are grateful to Courtney Myers, Laura Russo, Lindsay Smith, and Rebecca Trzupec for contributions to stimulus editing, recruitment, and data collection. Portions of this work were presented at the annual meeting of the American Auditory Society, Scottsdale, AZ (March 2013), and at the American Speech-Language-Hearing Association Annual Convention, Chicago, IL (November 2013).

References

- Allen, P., & Wightman, F. (1992). Spectral pattern discrimination by children. *Journal of Speech and Hearing Research*, 35, 225–235.
- Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, 14, 255–259.
- Aslin, R. N., & Smith, L. B. (1988). Perceptual development. *Annual Review of Psychology*, 39, 435–473.
- Bernstein, L. E., Auer, E. T., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication*, 44, 5–18.
- Binnie, C. A., Montgomery, A. A., & Jackson, P. L. (1974). Auditory and visual contributions to the perception of consonants. *Journal of Speech and Hearing Research*, 17, 619–630.
- Boothroyd, A. (1991). Speech perception measures and their role in the evaluation on hearing aid performance in a pediatric population. In J. A. Feigin & J. P. Stelmachowicz (Eds.), *Pediatric amplification* (pp. 77–91). Omaha, NE: Boys Town National Research Hospital.
- Boothroyd, A., Eisenberg, L. S., & Martinez, A. S. (2010). An On-Line Imitative Test of Speech Pattern Perception (OlimSpac): Developmental effects in normally hearing children. *Journal of Speech, Language, and Hearing Research*, 53, 531–542.
- Braida, L. D. (1991). Crossmodal integration in the identification of consonant segments. *The Quarterly Journal of Experimental Psychology*, 43A, 647–677.
- Burnham, D., & Dodd, B. (2004). Auditory-visual integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, 45, 204–220.
- Carney, A. E., Osberger, M. J., Carney, E., Robbins, A. M., Renshaw, J., & Miyamoto, R. T. (1993). A comparison of speech discrimination with cochlear implants and tactile aids. *The Journal of the Acoustical Society of America*, 94, 2036–2049.
- Coplan, J., & Gleason, J. R. (1988). Unclear speech: Recognition and significance of unintelligible speech in preschool children. *Pediatrics*, 82, 447–452.
- Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, 66, 85–110.
- Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, 45, 187–203.
- Dodd, B. (1977). The role of vision in the perception of speech. *Perception*, 6, 31–40.
- Eramudugolla, R., Henderson, R., & Mattingly, J. B. (2011). Effects of audio-visual integration on the detection of masked speech and non-speech sounds. *Brain and Cognition*, 75, 60–66.
- Erber, N. P. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research*, 12, 423–425.
- Erber, N. P. (1972). Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. *Journal of Speech and Hearing Research*, 15, 413–422.
- Erber, N. P. (1982). Glendonald auditory screening procedure. In N. P. Erber, *Auditory training* (pp. 47–71). Washington, DC: Alexander Graham Bell Association.
- Eskelund, K., Tuomainen, J., & Anderson, T. S. (2011). Multistage audiovisual integration of speech: Dissociating identification and detection. *Experimental Brain Research*, 208, 447–457.
- Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2010). The word superiority effect in audiovisual speech perception. *Speech Communication*, 52, 525–532.
- Fort, M., Spinelli, E., Savariaux, C., & Kandel, S. (2012). Audio-visual vowel monitoring and word superiority effect in children. *International Journal of Behavioral Development*, 36, 457–467.
- Gerken, L. (2002). Early sensitivity to linguistic form. *Annual Review of Language Acquisition*, 2, 1–36.
- Gershkoff-Stowe, L., & Thelen, E. (2004). U-shaped changes in behavior: A dynamic systems perspective. *Journal of Cognition and Development*, 5(1), 11–36.
- Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, 108, 1197–1208.
- Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *The Journal of the Acoustical Society of America*, 103, 2677–2690.
- Hillock-Dunn, A., & Wallace, M. T. (2012). Developmental changes in the multisensory temporal binding window persist into adolescence. *Developmental Science*, 15, 688–696.
- Hnath-Chisolm, T. E., Laipply, E., & Boothroyd, A. (1998). Age-related changes on a children's test of sensory-level speech

- perception capacity. *Journal of Speech, Language, and Hearing Research*, 41, 94–106.
- Hollich, G., Newman, R. S., & Jusczyk, P. W. (2005). Infants' use of synchronized visual information to separate streams of speech. *Child Development*, 76, 598–613.
- Holt, R. F. (2011). Enhanced speech discrimination through stimulus repetition. *Journal of Speech, Language, and Hearing Research*, 54, 1431–1447.
- Holt, R. F., & Carney, A. E. (2005). Multiple looks in speech sound discrimination in adults. *Journal of Speech, Language, and Hearing Research*, 48, 922–943.
- Holt, R. F., & Carney, A. E. (2007). Developmental effects of multiple looks in speech discrimination. *Journal of Speech, Language, and Hearing Research*, 50, 1404–1424.
- Holt, R. F., Kirk, K. I., & Hay-McCutcheon, M. J. (2011). Assessing multimodal spoken word-in-sentence recognition in children with normal hearing and children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 54, 632–657.
- Holt, R. F., & Lalonde, K. (2012). Assessing toddlers' speech discrimination. *International Journal of Pediatric Otorhinolaryngology*, 76, 680–692.
- Jerger, S., Damian, M. F., Spence, M. J., Tye-Murray, N., & Abdi, H. (2009). Developmental shifts in children's sensitivity to visual speech: A new multimodal picture word task. *Journal of Experimental Child Psychology*, 102, 40–59.
- Kirk, K. I., Pisoni, D. B., & Osberger, M. J. (1995). Lexical effects on spoken word recognition by pediatric cochlear implant users. *Ear and Hearing*, 16, 470–481.
- Kuhl, P. K., & Meltzoff, A. N. (1982, December 10). The bimodal perception of speech in infancy. *Science*, 218, 1138–1141.
- Kuhl, P. K., & Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, 7, 361–381.
- Lalonde, K., & Holt, R. F. (2014). Cognitive and linguistic sources of variance in 2-year-olds' speech-sound discrimination: A preliminary investigation. *Journal of Speech, Language, and Hearing Research*, 57, 308–326.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- MacLeod, A., & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, 21, 131–141.
- Markovitch, S., & Lewkowicz, D. J. (2004). U-shaped functions: Artifact or hallmark of development? *Journal of Cognition and Development*, 5, 113–118.
- Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development*, 55, 1777–1788.
- Massaro, D. W., Thompson, L. A., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology*, 41, 93–113.
- McGurk, H., & MacDonald, J. (1976, December 23). Hearing lips and seeing voices. *Nature*, 264, 746–748.
- Metsala, J. L., & Walley, A. C. (1988). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Hillsdale, NJ: Erlbaum.
- Morais, J., Bertelson, P., Cary, L., & Algeria, J. (1986). Literacy training and speech segmentation. *Cognition*, 24, 45–64.
- Munhall, K. G., & Vatikiotis-Bateson, E. (2004). Spatial and temporal constraints on audiovisual speech perception. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 177–188). Cambridge, MA: MIT Press.
- Osberger, M. J., Miyamoto, R. T., Zimmerman-Philips, S., Kemink, J., Stroer, B. S., Firszt, J. B., & Novak, M. A. (1991). Independent evaluation of the speech perception abilities of children with the Nucleus 22-Channel cochlear implant system. *Ear and Hearing*, 12(Suppl. 4), 66S–80S.
- Owens, E., & Blazek, B. (1985). Visemes observed by hearing-impaired and normal-hearing adult viewers. *Journal of Speech and Hearing Research*, 28, 381–393.
- Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, 22, 237–247.
- Patterson, M. L., & Werker, J. F. (2003). Two-month-olds match vowel information in the face and voice. *Developmental Science*, 6, 191–196.
- Patti Stripes Square Wave Grating Paddles [Apparatus and software]. (n.d.). La Salle, IL: Precision Vision. Retrieved from http://precision-vision.com/index.cfm/product/157_132/patti-stripes-square-wave-grating-paddles.cfm
- Psychology Software Tools. (2007). E-Prime 2.0 [Computer program]. Pittsburgh, PA: Author.
- Rosenblum, L. D. (2005). Primacy of multimodal speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 51–78). Malden, MA: Blackwell.
- Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception and Psychophysics*, 59, 347–357.
- Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multi-sensory speech perception continues into the late childhood years. *European Journal of Neuroscience*, 33, 2329–2337.
- Ross, L. A., Saint-Amour, D., Leavitt, V. M., Javitt, D. C., & Foxe, J. J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex*, 17, 1147–1153.
- Schwartz, J.-L., Berthommier, F., & Savariaux, C. (2004). Seeing to hear better: Evidence for early audio-visual interactions in speech identification. *Cognition*, 93, B69–B78.
- Sekiya, K., & Burnham, D. (2008). Impact of language on development of auditory/visual speech perception. *Developmental Science*, 11, 306–320.
- Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in Cognitive Neuroscience*, 7, 343–348.
- Studebaker, G. A. (1985). A rationalized arcsine transform. *Journal of Speech and Hearing Research*, 28, 455–462.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26, 212–215.
- Sussman, J. E., & Carney, A. E. (1989). Effects of transition length on the perception of stop consonants by children and adults. *Journal of Speech and Hearing Research*, 32, 151–160.
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributions to phonetic learning in 6-month-old infants. *Cognition*, 108, 850–855.
- Trehub, S. E., Schneider, B. A., & Henderson, J. L. (1995). Gap detection in infants, children. *The Journal of the Acoustical Society of America*, 98, 2532–2541.
- Tremblay, C., Champoux, F., Voss, P., Bacon, B. A., Lepore, F., & Théoret, H. (2007). Speech and non-speech audio-visual illusions: A developmental study. *PLoS ONE*, 2, e742.
- Tuomainen, J., Andersen, T. S., Tiippana, K., & Sams, M. (2005). Audio-visual speech perception is special. *Cognition*, 96, B13–B22.

-
- Tye-Murray, N., & Geers, A.** (2001). *Children's Audio-Visual Enhancement Test*. St. Louis, MO: Central Institute of the Deaf.
- Tye-Murray, N., Sommers, M. S., & Spehar, B.** (2007). Auditory and visual lexical neighborhoods in audiovisual speech perception. *Trends in Amplification, 11*, 233–241.
- van Wassenhove, V., Grant, K. W., & Poeppel, D.** (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia, 45*, 598–607.
- Vatakis, A., Ghazanfar, A. A., & Spence, C.** (2008). Facilitation of multisensory integration by the “unity effect” reveals that speech is special. *Journal of Vision, 8*(9), 1–11.
- Walley, A. C.** (2004). Speech perception in childhood. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 449–468). Malden, MA: Blackwell.
- Walley, A. C., Metsala, J. L., & Garlock, V. M.** (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing: An Interdisciplinary Journal, 16*, 5–20.
- Weiss, C. E., & Lillywhite, H. S.** (1976). *Communicative disorders*. St. Louis, MO: Mosby.
- Wightman, F., Allen, P., Dolan, T., Kistler, D., & Jamieson, D.** (1989). Temporal resolution in children. *Child Development, 60*, 611–624.
- Wightman, F., Kistler, D., & Brungart, D.** (2006). Informational masking of speech in children: Auditory-visual integration. *The Journal of the Acoustical Society of America, 119*, 3939–3949.
- Yehia, H., Rubin, P. E., & Vatikiotis-Bateson, E.** (1998). Quantitative association of vocal-tract and facial behavior. *Speech Communication, 26*, 23–43.
- Yoshida, K. A., Fennell, C. T., Swingle, D., & Werker, J. F.** (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12*, 412–418.
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E.** (2002). *Preschool Language Scale—Fourth Edition*. San Antonio, TX: The Psychological Corporation.